BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e135453
doi: 10.3897/biss.8.135453

OPEN ACCESS

Conference Abstract

BiotXplorer: Navigating Evidence-Based Biotic Interactions

Patrick Ruch[‡,§], Emilie Pasche[|], Pierre-André Michel[|], Déborah Caucheteur[‡,¶]

‡ HES-SO, Geneva, Switzerland
§ SIB Swiss Institute of Bioinformatics, Geneva, Switzerland
| SIB & HES-SO, Carouge, Switzerland
¶ SIB, Geneva, Switzerland

Corresponding author: Patrick Ruch (patrick.ruch@sib.swiss)

Received: 24 Aug 2024 | Published: 26 Aug 2024

Citation: Ruch P, Pasche E, Michel P-A, Caucheteur D (2024) BiotXplorer: Navigating Evidence-Based Biotic Interactions. Biodiversity Information Science and Standards 8: e135453. https://doi.org/10.3897/biss.8.135453

Abstract

BiotXplorer is an exploration tool to navigate biotic interactions in BiodiversityPMC (Pasche et al. 2023, Gobeill et al. 2020), a digitally native research library of articles for biomedical, biodiversity and environmental sciences stored in Journal Article Tag Suite (JATS)/BioC formats (Comeau et al. 2019). BiotXplorer pre-processes all documents and supplementary data thanks to the Swiss Institute of Bioinformatics (SIB) Literature Services (SIBiLS) to build pairs of species co-occurring in the same sentence together with a biotic interaction concept as defined in the Relation Ontology. A search service is built on top of this database, which aggregates all triplets matching the query and using taxonomic hierarchies to expand the search. Researchers can thus discover new biotic interactions and understand how they are supported by published evidence. We manually evaluated the precision of BiotXplorer with two benchmarks: 100 randomly selected biotic interactions from BiotXplorer and GLOBI (Global Biotic Interactions), a database of biotic interactions based on tabular datasets. Out of the 100 random triples generated by BiotXplorer, we achieved a precision of 31% when identifying the interacting species. For 74% of the correct interacting species, we accurately identified the type of interaction between the two species. The main causes of error were instances where passages listed multiple species, which can be automatically filtered out. For the second benchmark, we focused on a set of validated biotic interactions—instead of using potential ones—with 85% of the returned passages confirming an interaction between

© Ruch P et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

the two species. Our primary goal is to support the detection of biotic interactions across all species. While the precision is dependent on many factors, the vast amount of data it processes can reveal new insights and patterns. The inclusion of evidence for each triplet can support a wide range of One Health/Biosecurity (Hulme 2020) applications (e.g., eDNA characterization, virus spillover prediction). Furthermore, we are working on refining the system using different post-processing methods, such as reducing the volumes of triples by retaining only top-ranked, and therefore most reliable, triples.

Keywords

text mining, information retrieval, semantic annotations, publications

Presenting author

Patrick Ruch

Presented at

SPNHC-TDWG 2024

Acknowledgements

The SIB Literature Services (SIBiLS) received funding to develop Biodiversity PMC and the BiotXplorer from the European Union's Horizon 2020 Research and Innovation Action under grant agreement No 101007492 for the BICIKL project. The maintaining of the services are also supported by the ELIXIR Data Platform as well as by internal SIB grants

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Comeau DC, Wei C, Islamaj Dogan R, Lu Z, et al. (2019) PMC text mining subset in BioC: about three million full-text articles and growing. Bioinformatics 35 (18): 3533-3535. https://doi.org/10.1093/bioinformatics/btz070
- Gobeill J, Caucheteur D, Michel P, Mottin L, Pasche E, Ruch P (2020) SIB Literature Services: RESTful customizable search engines in biomedical literature, enriched with automatically mapped biomedical concepts. Nucleic Acids Research 48 https://doi.org/10.1093/nar/gkaa328

- Hulme P (2020) One Biosecurity: a unified concept to integrate human, animal, plant, and environmental health. Emerging Topics in Life Sciences 4 (5): 539-549. https://doi.org/10.1042/etls20200067
- Pasche E, Agosti D, Penev L, Groom Q, Flament A, Gobeill J, Ruch P (2023) Towards "Biodiversity PMC". Biodiversity Information Science and Standards 7 https://doi.org/10.3897/biss.7.111647